Python warm up

Segmentation data

```
#meta=>biosample_count=78
#plotpars=>title="Testing Custom Plot Parameters"
#plotpars=>subtitle="Some Chromosomes, Colors etc."
#plotpars=>chr2plot="3,5,7,8,11,13,16"
#plotpars=>color_var_dup_hex=#EE4500;color_var_del_hex=#09F911
#plotpars=>size_title_left_px=300
#plotpars=>size_text_title_left_px=10
#sample=>biosample_id=pgxbs-kftvhhmm;group_id=NCIT:C6393;group_label="Invasive D
#sample=>biosample_id=GSM252886;group_id=NCIT:C4504;group_label="Malignant Breas
biosample_id  chro  start stop  mean  variant_type  probes
pgxbs-kftvhhmm  1 911484  11993973  -0.4486 DEL .
pgxbs-kftvhhmm  1 12158755  22246766  0.2859 DUP  .
pgxbs-kftvhhmm  1 22346353  24149880  -0.5713 DEL .
pgxbs-kftvhhmm  1 24160170  33603123  0.0812  . .
pgxbs-kftvhhmm  1 33683474  37248987  -0.6478 DEL .
pgxbs-kftvhhmm  1 37391587  248655165 0.0342  . .
pgxbs-kftvhhmm  2 110819  240942225 -0.0007 . .
pgxbs-kftvhhmm  3 119131  4655519 -0.0122 . .
pgxbs-kftvhhmm  3 4662952 4857477 0.9273 DUP  .
```

https://docs.progenetix.org/file-formats/#pgxfreq-segment-cnv-frequencies

Python warm up

- Data link: https://progenetix.org/beacon/variants/?output=pgxseg&filters=NCIT:C3030

- Check the data first, and write your own script to access and download the data via python.

- Transfer the data to dataframe in pycharm, with proper columns.

Python warm up

- **Histplot**: You can start by exploring the data to understand its structure and distribution. For example, you can check the distribution of the ‘reference_name' values using a histogram
- **Count plot**: Count the number of CNV events per biosample
- **Heatmap of CNV Events:** If you want to explore relationships between biosamples and CNV events, you can create a heatmap to visualize the presence or absence of CNV events across biosamples.

https://seaborn.pydata.org/generated/seaborn.histplot.html

https://doi.org/10.1093/database/baab043

- What is CNV/CNA?
- How will you describe or introduce progenetix (scale, data source, cancer types and so on)?
- Describe NCIt, ICOD, UBERON codes, and their relationships.
- What are CNV segmentations and CNV frequencies, and how to use them?
- What are APIs and how to use APIs in progenetix?
- How does progenetix visualise CNA profiles?
- What do you think should be improved in progenetix?

Please upload your file to your own folder of Bio392 GitHub, and name the file as lastname_firstname_paper_reading_day2.md.
It will be graded.

https://progenetix.org/

https://docs.github.com/en/get-started/writing-on-github/getting-started-with-writing-and-formatting-on-github/basic-writing-and-formatting-syntax